(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) EP 1 333 371 A2

# (12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
**06.08.2003 Bulletin 2003/32**

(51) Int Cl.[7]: **G06F 3/06**

(21) Application number: **03008580.7**

(22) Date of filing: **09.01.1997**

(84) Designated Contracting States:
**DE FR GB**

(30) Priority: **10.01.1996 JP 205096**

(62) Document number(s) of the earlier application(s) in accordance with Art. 76 EPC:
**97100286.0 / 0 784 260**

(71) Applicant: **Hitachi, Ltd.**
**Chiyoda-ku, Tokyo (JP)**

(72) Inventors:
- **Ozawa, Koji**
  **Odawara-shi, Kanagawa-ken (JP)**
- **Sano, Kazuhide**
  **Odawara-shi, Kanagawa-ken (JP)**
- **Koide, Takeshi**
  **Odawara-shi, Kanagawa-ken (JP)**
- **Nakamura, Katsunori**
  **Odawara-shi, Kanagawa-ken (JP)**

(74) Representative: **Strehl Schübel-Hopf & Partner**
**Maximilianstrasse 54**
**80538 München (DE)**

Remarks:
This application was filed on 14 - 04 - 2003 as a divisional application to the application mentioned under INID code 62.

## (54) Storage system

(57) In a data processing system in which main and sub disk storage devices 5, 10 are each under the control of individual disk control devices 3, 8, the write processing time is reduced by selectively sending data according to a command-chaining time between the main and sub disk control devices. A section 36 for judging cable length and function of the sub disk control device 8 estimates the command-chaining time between a pair of main and sub disk storage devices. A channel command analysing section 31 estimates the number of records to be transferred and the length of a record using a LOCATE RECORD command. The command judgement section 32 for the sub disk control device 8 optimises the command chain to be issued to the sub disk control device using the above information. A command issuing section 35 then issues the optimised command chain to the sub disk control device 8. Thus, a shorter transmission time is realised by sending either individual records or an entire track of data.

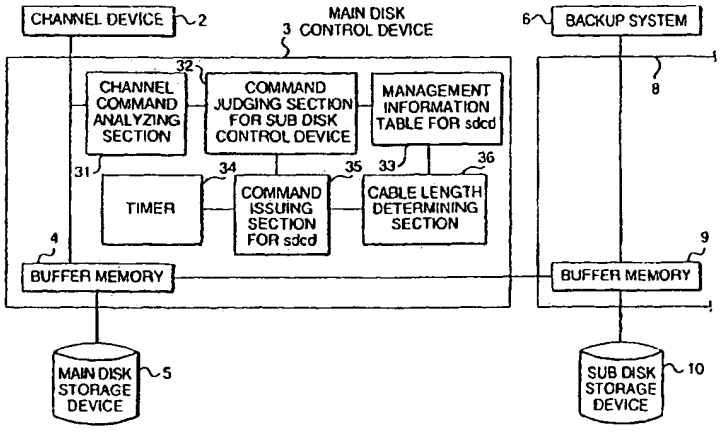


*FIG. 2*

EP 1 333 371 A2

## Description

### FIELD OF THE INVENTION

**[0001]** Present invention relates to an external storage control device which stores data to an external storage device according to a write command from a host. In particular, the present invention relates to a system in which the same data is copied to another external storage device.

### BACKGROUND OF THE INVENTION

**[0002]** When data used in a host are stored to plural external storage devices, main and sub external storage devices that hold the same data are sometimes provided under individual eternal storage control devices. In this case, these external storage control devices are mutually connected, and the main external storage control device issues a write command to the external storage control device which controls sub external storage devices when the external storage control device which controls main external storage devices receives a write command. Thus, data stored in the main and sub external storage devices are duplicated.

**[0003]** US 5,155,845 A discloses a method in which an external storage control device which controls main external storage devices and one which controls sub external storage devices are mutually connected. A main external storage control device which has received a write command from a host, transfers data to a sub external storage control device. Thus, the write process in performed in parallel for both of main and sub external storage devices.

**[0004]** DE 39 35 235 A discloses a storing method for microcomputer systems in which data are transferred between a central processor and a storage medium at high transfer rates by concatenating data to from composite packets, thereby minimising transmission time.

**[0005]** WO 94/19748 discloses a method of transferring data in blocks among computer input/output devices with the blocks being continually re-sized during transfer.

**[0006]** EP-A-0 601 738 describes a system in which data blocks are buffered during their transfer to a disc memory to save transfer time.

**[0007]** When a host handles data stored in external storage devices with CKD (count, key, data) format as used in large scale computer systems, the host issues channel commands in succession for instructing data transfer of each individual record. Thus, each individual record undergoes the same command chaining sequence in order to be transferred.

**[0008]** Figure 3 illustrates a case of plural records being written using the same command chaining sequence. This Figure illustrates the case in which plural records are written in succession by a single command-chaining. In an external storage sub system having main and sub external storage devices that hold the same data, located under individual external storage control devices, when a channel device that is a host issues a command-chain (DEFINE EXTENT / LOCATE RECORD / WRITE(R1) / WRITE(R2) / WRITE(R3)) to write 3 successive records R1, R2, and R3 to a disk storage device that is a main external storage control device, data flow between the channel device and the main external storage control device and between the main external storage control device and the sub external storage control device are shown in the processing sequences of Figure 3. Thus, several command-chains between the main and sub external storage control devices for each data transfer of write records is executed.

**[0009]** Further, in case data is duplicated by adopting the system described above, the distance between main and sub external storage control devices becomes large considering the backups necessary in case of a disaster. Thus, an optical fibre cable is adopted as the interface cable. Accordingly, the influence of cable delay, which is considered to be constant for a metal cable, with respect to command-chaining time cannot be ignored. However, since this duplication is for back up purposes, the influence of the write process for the sub external storage control device during an ordinary process must be minimised.

**[0010]** In a case where the write command must be executed a number of times in accordance with the command-chaining described above, this command-chaining time required for writing to sub external storage control devices cannot be ignored. This is because the amount of command-chaining between main and sub external storage control devices increases. As a result, backup processing can severely decrease the throughput rate in ordinary processing.

### SUMMARY OF THE INVENTION

**[0011]** Thus, one purpose of present invention is to optimise the write process time for the sub external storage control device by taking into consideration the command-chaining time between main and sub external storage control devices. As a result, the present invention offers a means for achieving excellent performance even under the conditions explained above.

**[0012]** In order to achieve the above-mentioned purpose, the external storage control device according to present invention is equipped with a means for estimating command-chaining time between main and sub external storage control devices. There is also provided a means for estimating the time for a write process to a sub external storage control device before starting the write process to the sub external storage control device. The command is issued to the main external storage control device from the host. The present invention also includes means to select the best suited command-chaining method, Thus, a comparison is made between the above-mentioned estimated time with the data

processing time required in the case of transferring data for a write record or data of plural tracks including a write record in one operation using a specified command. The present invention also includes command means for writing said data in a single operation.

**[0013]** It is possible for an external storage device to learn of the command-chaining time mentioned above by either measuring the command-chaining time from a specified command to the next command, or by setting the length of interface cable between main and sub external storage control devices from outside in advance. On the other hand, the time required for a write process for the sub external storage control device can be calculated by the above-mentioned information and by the information included in the command issued to the main external storage control device. Namely, command-chaining used for data input/output includes at least two specific channel commands prior the command to start data transfer. For example, for an external storage control device, the commands, "DEFINE EXTENT" and "LOCATE RECORD", are issued prior to data transfer, and the number of records to be processed and the data length are given so that the amount of data to be transferred can be calculated.

**[0014]** Furthermore, the command-chaining time between main and sub external storage control devices depends upon length of interface cable and the performance of the external storage control device, and does not depend on commands made before and after the command-chaining. Accordingly, command-chaining time during data transfer can be estimated by measuring the time for two command chains. Also, the length of interface cable between main and sub external storage control devices, and support functions of a given external storage control device are known at the time of installation of a backup system. Thus, command-chaining time can be estimated with the length of interface cable as established.

**[0015]** Since the command-chaining time can be estimated as mentioned above, the processing time can also be estimated for a given amount of data transfer in the case that the command-chaining command from the channel device is issued to the sub external storage control device. Consequently, by comparing this processing time with a processing time in the case of transferring data for a write record, or, data of plural tracks including a write record together, using a command and command means to write said data at once, the best suited command-chaining instruction can be issued to the sub external storage control device.

**[0016]** Also, the external storage control device according to present invention is equipped with means to confirm that the object data exists in the data buffer of another external storage control device in order to transfer data from the data buffer within the external storage control device to an external storage control device that receives the write commands, in addition to data transfer means as mentioned above. Moreover, the processing time can be further reduced by transferring all the physical data together, including the control byte and the check byte in the data buffer.

**[0017]** Additionally, in case where object data for the read command, received from the host, exists in the buffer of an external storage control device other than the external storage control device which has received the read command, the processing time for performing a read-out can be shortened by transferring these data to the external storage control device which received the read command instead of accessing the other external storage device.

**[0018]** These and other objects, features and advantages of the present invention will become more apparent in view of the following detailed description of the preferred embodiments.

## BRIEF DESCRIPTION OF THE DRAWINGS

**[0019]**

Figure 1 illustrates a configuration of the data processing system according to an embodiment of the present invention.

Figure 2 illustrates a configuration of the main disk control device according to an embodiment of the present invention.

Figure 3 illustrates the data flow between the channel device, that is a host, the main disk control device, and the sub disk control device.

Figure 4 illustrates the data flow in the case that data are transferred between the main and sub disk control devices record by record or in an entire track.

Figure 5 illustrates data format in the data buffer of a disk control device for the case of 8 records per track.

Figure 6 illustrates a graph showing the relation between number of records transferred and the processing time.

Figure 7 illustrates a flow chart representing steps followed according to an embodiment of the present invention.

## DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

**[0020]** The preferred embodiments of the present invention will now be described in conjunction with the Figures. Figure 1 illustrates a configuration diagram of data processing system according to an embodiment of the present invention. This data processing system includes a main memory device 1, a channel device 2, a backup system 6 and a channel device 7. This data processing system also includes a main disk control device 3 equipped with a buffer memory 4, and a main disk storage device 5, and also includes a sub disk control device 8 equipped with a buffer memory 9, and a sub disk stor-

age device 4. The channel device 2 is connected to the main disk storage control device 3 through an interface cable 11, the main disk storage control device 3 is connected to the sub disk storage control device 8 through an interface cable 12. Interface cables are also used to interconnect other elements shown in this Figure. Preferably, optical fibre cables are used as the interface cables.

**[0021]** As shown in Fig. 2 the main disk control device 3 includes a channel command analysing section 31, a command judging section for judging commands to be issued to the sub disk control device 32, a management information table for the sub disk control device 33, a timer 34, a command issuing section for the sub disk control device 35, and a section for interpreting cable length and the function of the sub disk control device 36. Function blocks 31 33, 35 36 are realised by a microprogram executed by a microprocessor contained in the disk control device 3. Timer 34 is a hardware counter that is counted up according to a constant period. The sub disk control device 8 is constructed similarly to main disk control device 3.

**[0022]** The command-chaining time required between main and sub external storage control devices is measured in the main disk control device 3, by using the timer 34, the section for issuing command to the sub disk control device 35, and the section for interpreting cable length and function of the sub disk control device 36. Then, the command chaining time, Tsg, is measured when a command is processed once a pair of main and sub disk storage devices are established. The measuring is performed by timer 34 and is stored in the management information table 33 in the main disk control device 3.

**[0023]** The main disk control device obtains support function level information of the sub disk control device when the pair of main and sub disk storage device is established. This function level information is stored in the management information table of the main disk control device. This stored information is used to judge whether a write command to write data for write record, or, data of plural tracks including the write record can be accepted by sub disk storage control device or not.

**[0024]** If command judgement section for the sub disk control device 32 has judged from the support function level information that a specified command is not acceptable, when a write command is received from the channel device 2, the section for issuing command to the sub disk control device 35 unconditionally issues a command-chain sent from the channel device to the sub disk control device. Thus, the system is rendered more versatile in that it can connect even with a conventional device that does not support these functions.

**[0025]** The channel command analysing section 31 operates as follows. When the channel device 2 issues a write command to the main disk control device 3, the channel command analysing section 31 determines whether or not the write command is for a disk storage device that forms a pair with the main disk control device 3.

**[0026]** Figure 6 is a graph illustrating command-chaining time for two separate cases. One case corresponds to the processing time required when data for one track including a write record. The information from the management information table for the sub disk control device 33 is used, as is the number of transfer records of LOCATE command, the number N which is analysed by the channel command analysing section 31, and average record length, L. Here, the straight line (II) can be explained by an equation as shown below using known values of command chaining time Tsg between the main and sub external storage control device, the above analysis, and the data transfer velocity V between the main and sub external storage control devices 3 and 8.

$$Ts = NL/V+(N-1)Tsg \quad (1)$$

**[0027]** The straight line (II) shows the transfer time Ts', which is a constant value, for the case where data of one track is transferred together and where each record length is L, and number of records in a track is Na. Ts' can be represented by the following formula.:

$$Ts' = NaL/V+Tsg \quad (2)$$

**[0028]** In the case of the sloped line (I), as the number of records increases, the processing time increases. At some point, i.e. for some number of records, it is faster to send the entire track of data, because the command chaining time Tsg is too great.

**[0029]** The section for judging commands to be issued to the sub disk control device 32 judges whether to issue the command chain to the sub disk control device as received from the channel device or to transfer data of one track all together based upon the information contained in Figure 6, in view of the number of records N.

**[0030]** Figures 5 and 6 illustrate a case in which six records are to be written in a track that holds eight records of data. As seen from Figure 4, transferring data of one track all together requires less processing time if the number of record is greater than 4. Figure 6 also illustrates such a relationship according to the intersection of lines I and II. Command issuing section 35 issues the best suited command-chain to the sub disk control device 8, in view of all of the considerations mentioned above and according to the instruction of command judgement section 32.

**[0031]** Figure 3 illustrates command chaining between a CPU (channel device or host) and a sub disk controller via a main disk controller. As shown, a number of commands have to be sent back and forth before a record (R1) can be transmitted. The same sequences

is followed for subsequent records.

**[0032]** According to the present invention, write processing time to a sub disk storage device, i.e. back-up processing time, can be minimised by selecting a command-chain method to be issued to the sub disk control device. This is accomplished by comparing processing time for the case of issuing the command chain to the sub disk control device as received from the channel device with that of the case of transferring data for a write record, or, data of plural tracks including a write record all together.

**[0033]** Figure 7 illustrates the flow of steps according to the present invention. First, the amount of data that is to be sent is calculated (STEP 100), i.e. the number of records. Then a time T1 is calculated (STEP 110). T1 is equal to Ts using equation (1) above, and is computed for the case of sending individual records. Then, a time T is calculated (STEP 120). T2 is equal to Ts using equation (2) above, and is computed for the case of sending the entire track of data. In decision block 130, it is determined if T1>T2 (STEP 140). If so, then the entire track of data is sent (STEP 140). If not, then the individual records are sent separately (STEP 150). This process is repeated as necessary.

**Claims**

1. A storage system comprising:

   a first plurality of disk drives (5) and
   a first disk controller (3) connected to a host processor (2) and said first plurality of disk drives (5) for controlling the data transfer between said host processor (2) and said first plurality of disk drives (5),

   wherein said first disk controller (3) is connected to a second disk controller (8) and obtains information on a support function of the second disk controller (8).

2. The system of claim 1, wherein said first disk controller (3) issues to the second disk controller (8) information that is different according to the support function of the second disk controller (8).

3. The system of claim 2, wherein said first disk controller (3) acts according to a first support function of the second disk controller (8) if the second disk controller (8) has the first support function.

4. The system of claim 2, wherein said first disk controller (3) issues to the second disk controller (8) information that does not correspond to second support function, if the second disk controller (8) does not have the second support function.

5. The system of claim 2, wherein said first disk controller (3) includes a function determining section for judging the support function of the second disk controller (8).

6. The system of claim 1 or 5, wherein said first disk controller (3) includes a management table for storing the information on the support function of the second disk controller (8).

7. The system of claim 1 or 6, wherein said first disk controller (3) further includes means for obtaining the information on the support function of the second disk controller (8) from the second disk controller (8).

8. The system of claim 7,

   wherein said first disk controller (3) establishes a pair of said first plurality of disk drives (5) and a second plurality of disk drives (10) controlled by the second disk controller (8), and

   wherein said first disk controller (3) obtains the information on the support function of the second disk controller (8) from the second disk controller (8) when the pair of said first and second disk drives (5, 10) is established.

9. The system of claim 7, wherein said first disk controller (3) stores in said management table the information on the support function of the second disk controller (8) and obtained from the second disk controller (8).

10. The system of claim 9 as dependent on claim 5, wherein said first disk controller (3) uses the information on the support function of the second disk controller (8) in said management table so that said function determining section judges the support function of the second disk controller (8).

11. The system of claim 2, wherein said first disk controller (3) does not issue a specified command to the second disk controller (8), if said first disk controller (3) judges from the information on the support function of the second disk controller (8) that the second disk controller (8) cannot accept the specified command.

12. A storage system comprising:

    a first plurality of disk drives (5) and
    a first disk controller (3) connected to host processor (2) and said first plurality of disk drives (5) for controlling the data transfer from said first plurality of disk drives (5) to said host processor (2),

    wherein said first disk controller (3) is con-

nected to another storage system and transfers a read command received from the host processor (2) to the other storage system, in the case where object data for the read command exist in the other storage system.

**13.** The system of claim 12,

wherein said first disk controller (3) controls the writing of data from said host processor (2) to said first plurality of disk drives (5), and

wherein said first disk controller (3) includes means for confirming that object data for the read command or a write command received from the host processor (2) exist in the other storage system, and means for transferring the read or write command from first disk controller (3) to the other storage system.

**14.** A storage system comprising:

a first plurality of disk drives (5), and
a first disk controller (3) connected to said first plurality of disk drives (5) for controlling the data transfer between said host processor (2) and said first plurality of disk drives (5),

wherein said first disk controller (3) is connected to a second disk controller (8) which supports another function against said first disk controller (3).

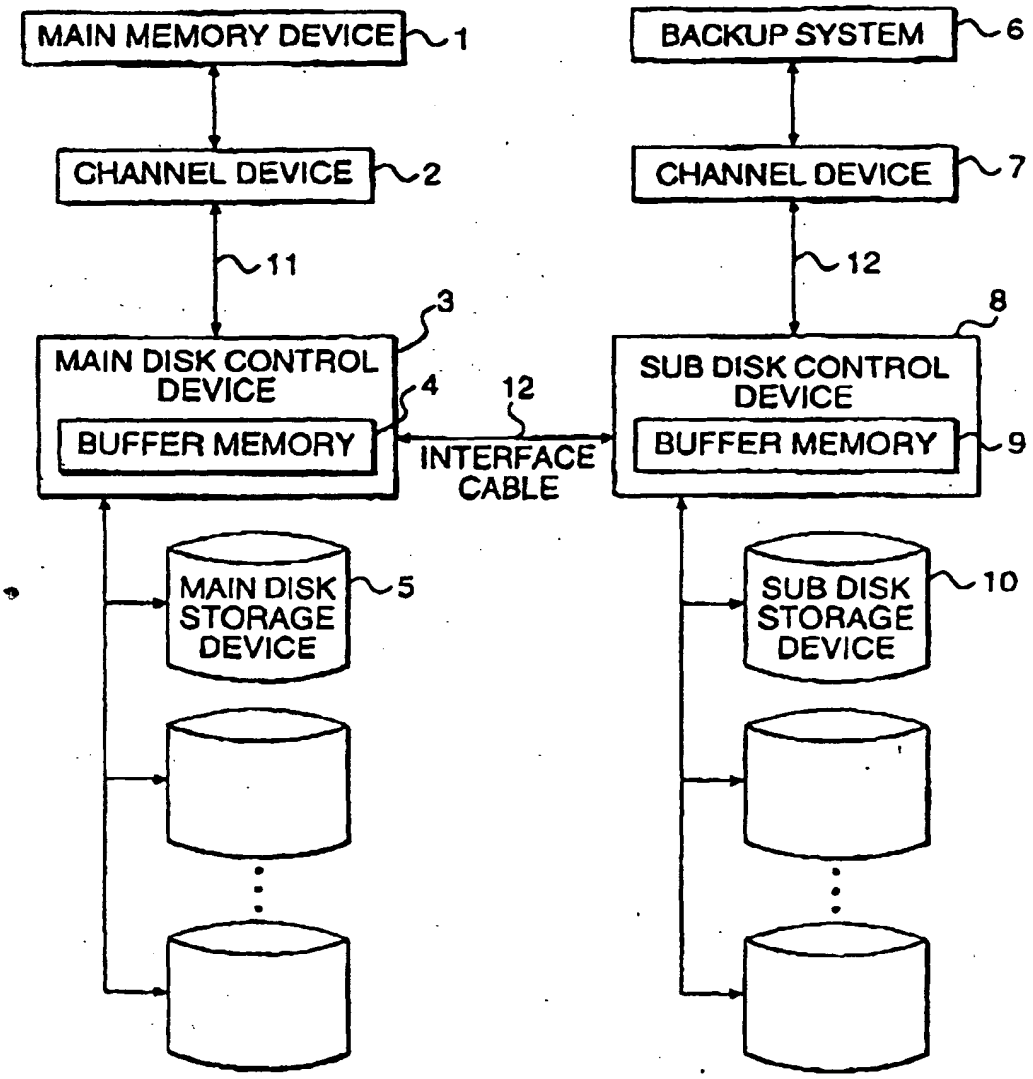
MAIN MEMORY DEVICE
1
BACKUP SYSTEM
6
CHANNEL DEVICE
2
CHANNEL DEVICE
7
11
12
3
8
MAIN DISK CONTROL DEVICE
4
12
SUB DISK CONTROL DEVICE
BUFFER MEMORY
INTERFACE CABLE
BUFFER MEMORY
9
MAIN DISK STORAGE DEVICE
5
SUB DISK STORAGE DEVICE
10

FIG. 1

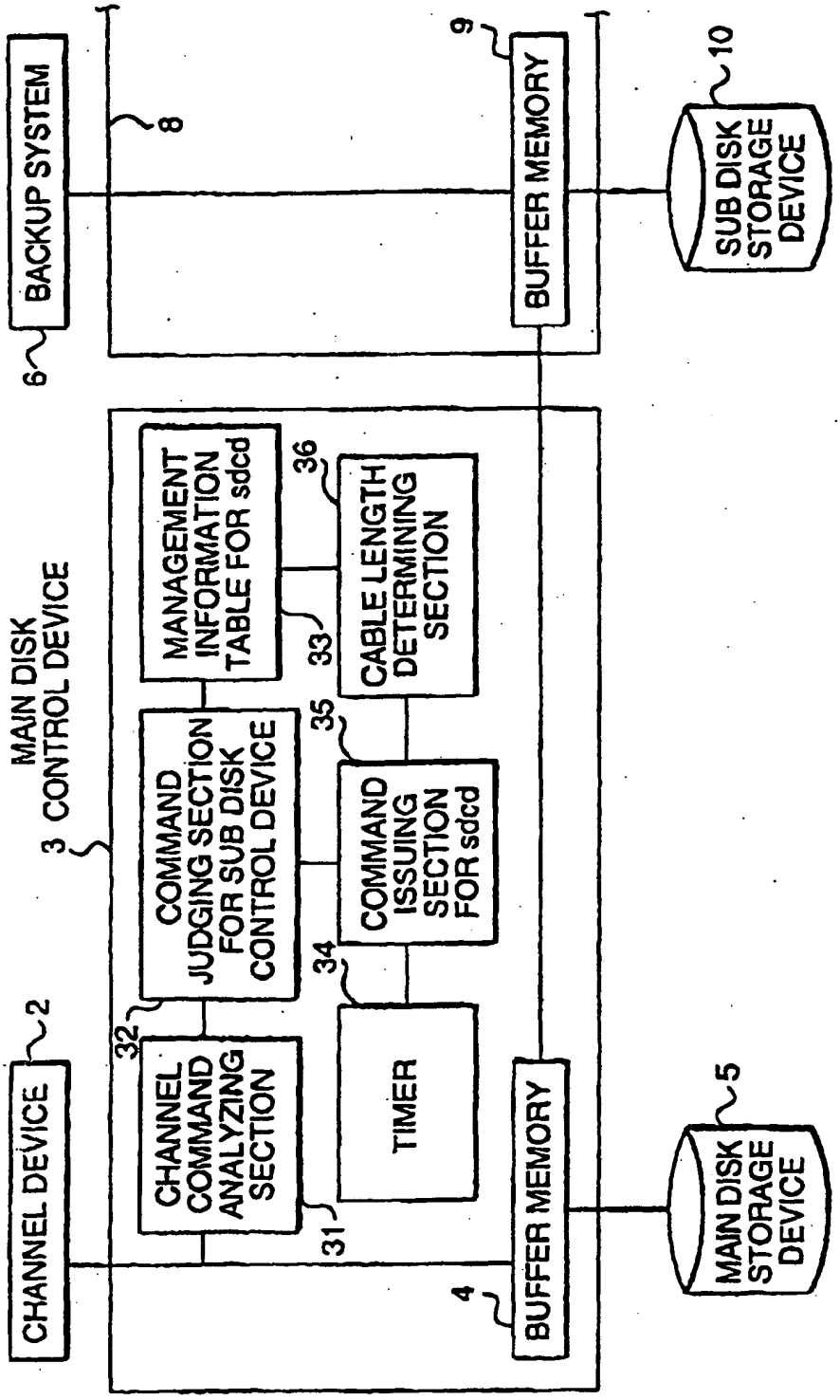
BACKUP SYSTEM
6
8
9
BUFFER MEMORY
10
SUB DISK STORAGE DEVICE
MAIN DISK CONTROL DEVICE
3
MANAGEMENT INFORMATION TABLE FOR sdcd
36
CABLE LENGTH DETERMINING SECTION
33
COMMAND JUDGING SECTION FOR SUB DISK CONTROL DEVICE
35
COMMAND ISSUING SECTION FOR sdcd
34
CHANNEL DEVICE
2
32
CHANNEL COMMAND ANALYZING SECTION
TIMER
31
4
BUFFER MEMORY
5
MAIN DISK STORAGE DEVICE

FIG. 2

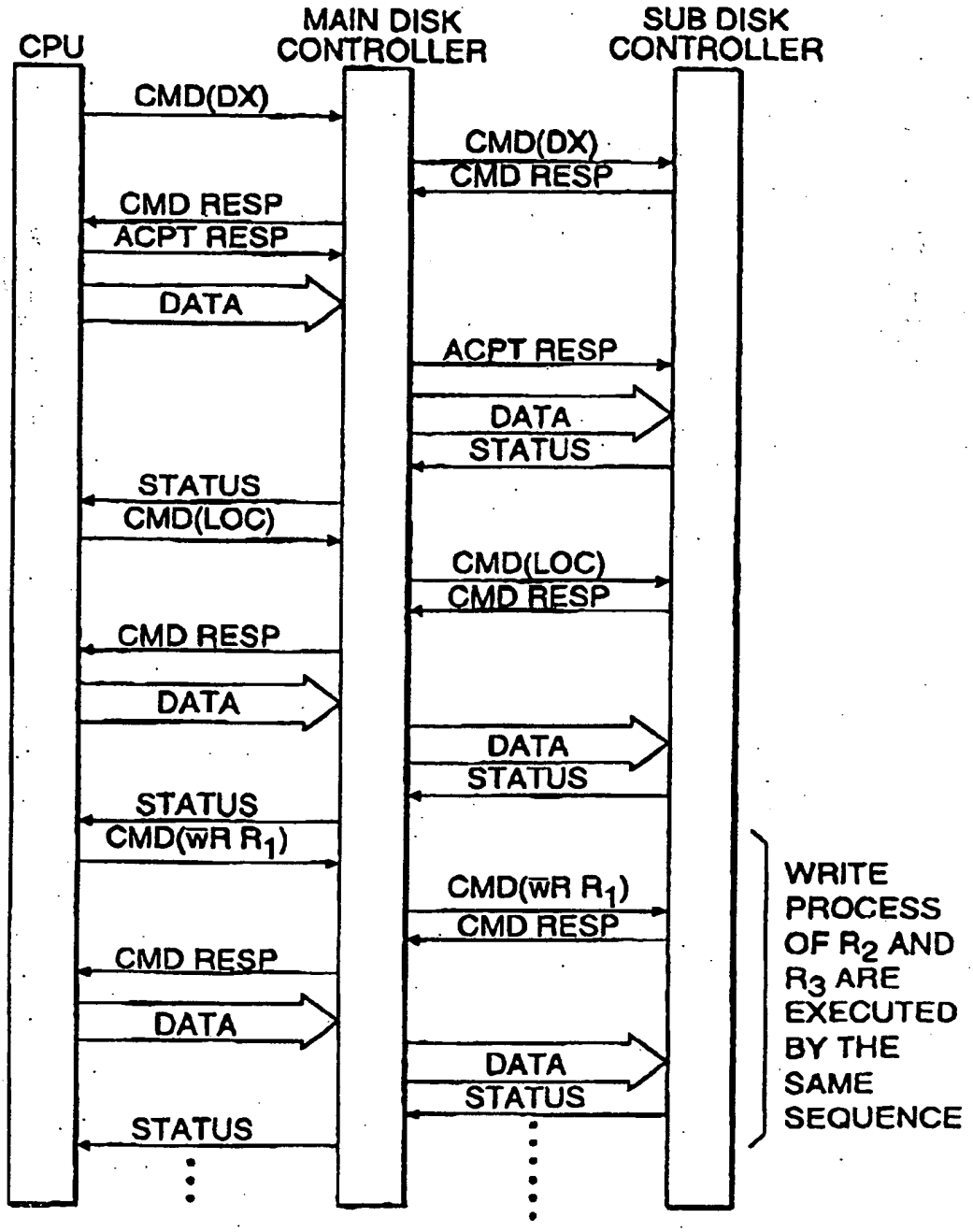
CPU
MAIN DISK CONTROLLER
SUB DISK CONTROLLER
CMD(DX)
CMD(DX)
CMD RESP
CMD RESP
ACPT RESP
DATA
ACPT RESP
DATA
STATUS
STATUS
CMD(LOC)
CMD(LOC)
CMD RESP
CMD RESP
DATA
DATA
STATUS
STATUS
CMD(w̄R R1)
CMD(w̄R R1)
CMD RESP
CMD RESP
DATA
DATA
STATUS
STATUS
WRITE PROCESS OF R2 AND R3 ARE EXECUTED BY THE SAME SEQUENCE

FIG. 3

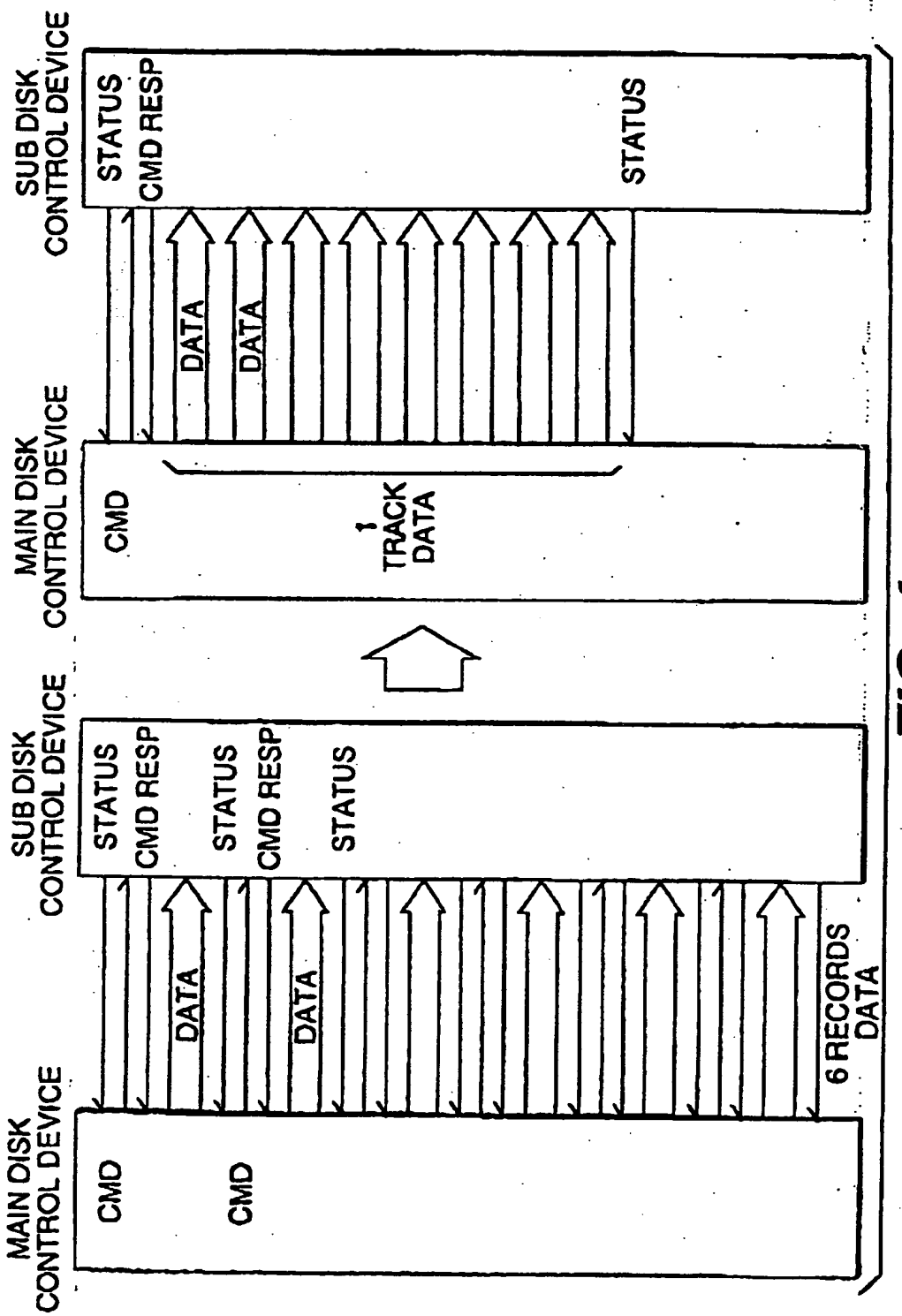
MAIN DISK CONTROL DEVICE
CMD
CMD
SUB DISK CONTROL DEVICE
STATUS
CMD RESP
DATA
STATUS
CMD RESP
DATA
STATUS
6 RECORDS DATA
MAIN DISK CONTROL DEVICE
CMD
1 TRACK DATA
SUB DISK CONTROL DEVICE
STATUS
CMD RESP
DATA
DATA
STATUS

FIG. 4

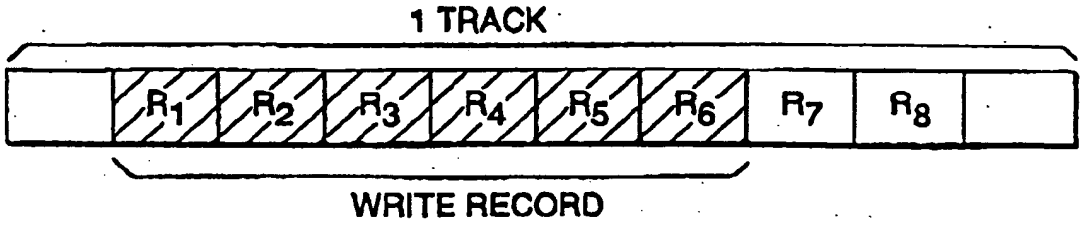
1 TRACK
R1
R2
R3
R4
R5
R6
R7
R8
WRITE RECORD

FIG. 5

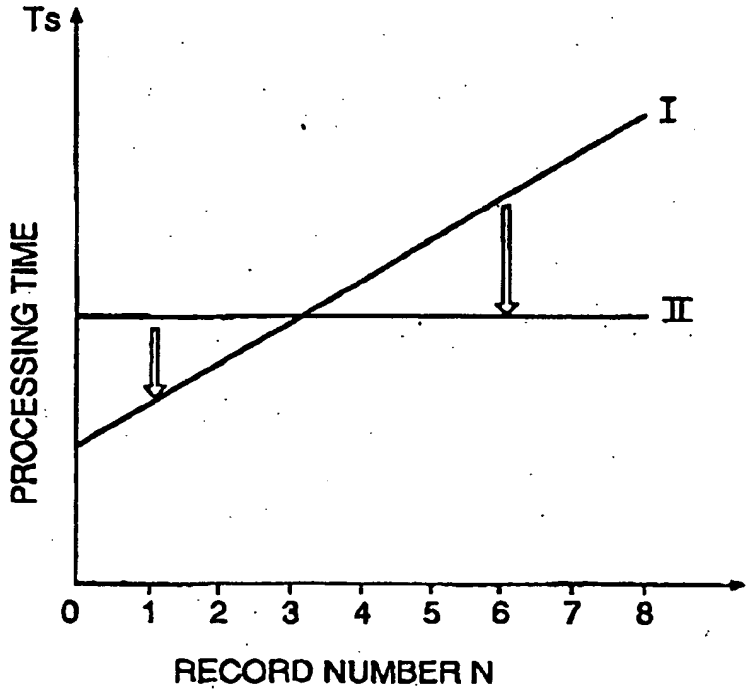
Ts
I
II
PROCESSING TIME
0
1
2
3
4
5
6
7
8
RECORD NUMBER N

FIG. 6

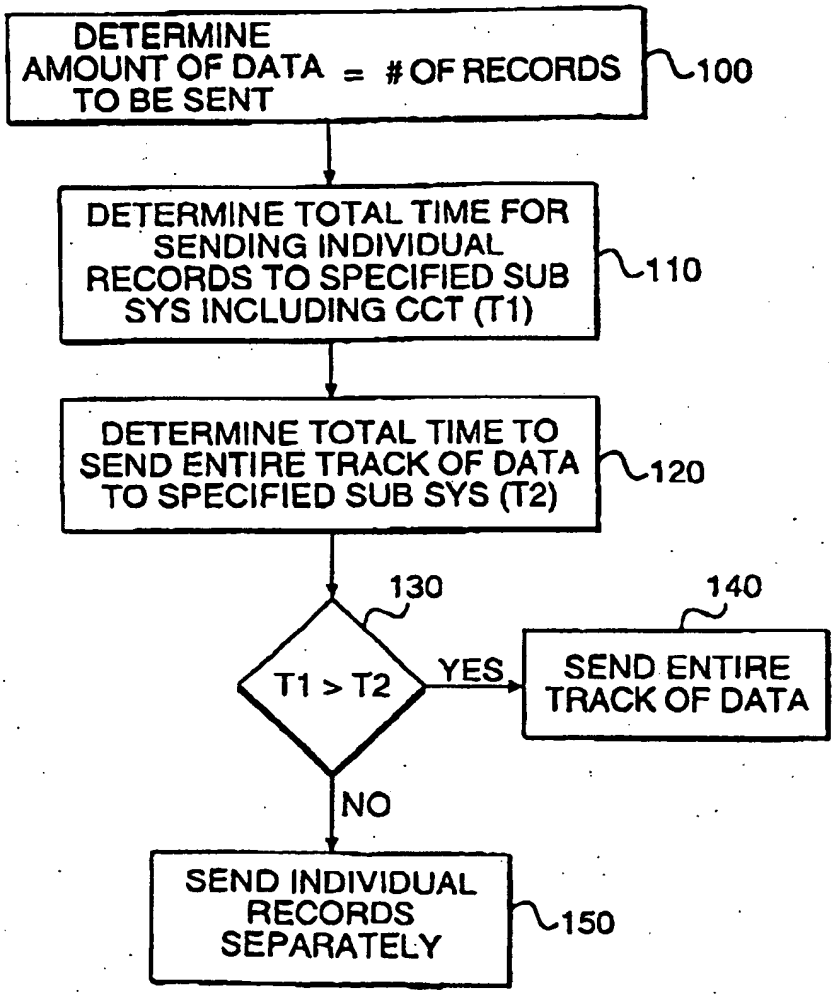
DETERMINE AMOUNT OF DATA TO BE SENT = # OF RECORDS
100
DETERMINE TOTAL TIME FOR SENDING INDIVIDUAL RECORDS TO SPECIFIED SUB SYS INCLUDING CCT (T1)
110
DETERMINE TOTAL TIME TO SEND ENTIRE TRACK OF DATA TO SPECIFIED SUB SYS (T2)
120
130
T1 > T2
YES
140
SEND ENTIRE TRACK OF DATA
NO
SEND INDIVIDUAL RECORDS SEPARATELY
150

FIG. 7